

# A Comparative study of Usability Evaluation Methods

Sugandha Gupta<sup>#1</sup>

<sup>#</sup>Assistant Professor, St. Stephen's College, Department of Computer Science, University of Delhi

**Abstract**— Usability is an essential quality of software systems. Many techniques have so far been projected for usability evaluation but they are not well integrated and fail to ensure all the aspects of usability. The inception of usability evaluation methods (UEMs) to assess and improve usability in such systems has led to a variety of alternative methodologies and a general lack of understanding of the proficiencies and limitations of each. This misperception has exaggerated the need for practitioners and others to regulate which methods are more operative, and in what ways and under what scenarios. However, UEMs cannot be assessed and equated reliably because of the lack of standard criteria for comparison. This paper aims at discussing various existing usability evaluation methods (UEM) that have been presented till 2015.

**Keywords**— Usability, Usability Evaluation Methods (UEMs), Inspection Methods, Human Computer Interaction, Inquiry Methods, Testing Methods.

## I. INTRODUCTION

Usability is evaluated by the quality of communication (interaction) between a technological product (system) and a user (the one who uses that technological product). In other words, usability evaluation reaches back to virtually the beginning of human-computer interaction (HCI). However, in an extensive historical view, the field is still comparatively new and partially complete as both a research topic and as an applied body of knowledge.

Till date, various usability evaluation methods (UEMs) have been proposed by the research practitioners. These can be classified as inspection, testing and inquired methods. Inspection methods focuses on user interface, testing methods focuses on task performance and inquired methods focuses on user data. Assortment of one these evaluation methods are grounded on number of parameters like available resources, abilities of evaluator, types of users, environment etc.

The paper is aiming to provide a literature survey of all the available methodologies under the three categories of UEMs in the subsequent sections. The rest of the paper is organized as follows: Section I discusses various Inspection methods, Section II discusses various Testing methods and Section III discusses various Inquiry methods. Finally, Section IV concludes this paper.

## II. INSPECTION METHODS

These techniques comprises of a set of methods that are all based on having evaluators inspect a user interface with

respect to its conformance to a set of guidelines. Guidelines can range from highly specific prescriptions to broad principles

### A. Cognitive Walkthrough

Cognitive walkthrough (Lewis et al., 1990; Wharton et al., 1992; Rieman, Franzke and sRedmiles, 1995) is a theoretically structured usability evaluation process that focuses on a user's cognitive activities, especially while performing a task. Cognitive walk through involves one or more evaluators discovering an interface, prototype, or paper mock-up by going through a pre-determined set of tasks and measuring the understandability and easiness of learning for each task.

It lays the emphasis on user and helps in identifying the user goals.

### B. Heuristic Evaluation

Heuristic evaluation is the most informal inspection method [Nielsen and Mack 1994], mainly because it count upon a small set of usability criteria. In this technique, one or more evaluators autonomously evaluate an interface using a list of heuristics. HCI experts separately review an interface and categorise and justify problems based on a short set of heuristics. The outcome of this evaluation is typically a list of possible usability problems.

### C. Feature Inspection

The purpose of this evaluation method [Nielsen (1994)] is to inspect a feature set of a product and to analyse the availability, understandability, and other functionality aspects for each feature. Evaluators use a list of product features along with situations for such inspections.

### D. Pluralistic Walkthrough

Pluralistic Walkthrough [Bias (1994)] is a variation of the cognitive walk through inspection method wherein representative users, evaluators, and developers inspect the interface as a group.

### E. Perspective based Inspection

Perspective-based inspection [Zhang 1998] is a variation of heuristic evaluation. Interfaces are inspected from three diverse perspectives i.e. novice use, expert use and error handling; considering one perspective at a time.

### F. Formal Usability Inspection

It is a six step procedure that combines heuristic evaluation and cognitive walkthrough. The steps include planning, kick-

off meeting, review, logging meeting, rework and follow-up [Bell (1992)].

#### *G. Consistency Inspection*

Evaluators use this method to conclude a consistent interface appearance and functionality that they can then use to weigh the uniformity of interfaces across multiple products in a family. It gives a summary of the inconsistencies [Wixon et al. (1994)].

#### *H. Standards Inspection*

In this inspection method [Wixon et al. (1994)], an evaluator equates components of an interface to a list of industry standards to assess the interface's compliance with these standards. This inspection method is usually aimed at ensuring a product's market conformance.

### III. TESTING METHODS

These techniques are the best way to understand how real users experience a particular software. During usability testing, participants use the system or a prototype to complete a pre-determined set of tasks while the tester or software records the results of the participants' work.

#### *A. Remote Testing*

In this method [Hartson et al. (1996)], the testers and participants are separated in space and/or time. It may be same time different place or different time different place, depending on the need.

Remote usability testing is used when tester(s) are disconnected in space and/or time from the participants. This means that the tester(s) cannot observe the testing process directly and that the participants are usually not in a formal usability laboratory. There are different types of remote testing. One is same-time but different-place, where the tester can observe the test user's screen through computer network, and may be able to hear what the test user says during the test through speaker telephone. Another is different-time different-place testing, where the user's test session is guided and logged through a special piece of software as well as additional code added to the system being tested

#### *B. Coaching Method*

The coaching method [Nielsen (1993)] allows participants to ask any system related questions to an expert during usability testing. The main goal of this method is to define the information needs of users to deliver improved training and documentation in addition to probably redesigning the interface to eradicate the need for questions in the first place.

The purpose of this procedure is to determine the information requirements of users in order to provide improved training and documentation, as well as possibly redesign the interface to evade the need for the questions.

#### *C. Performance Measurement*

The goal of this testing method [Nielsen (1993)] is to capture quantitative data about participants' performance when they complete tasks. As such, there is typically no collaboration between the tester and participant during the test.

This technique is used to acquire quantitative data about test participants' performance when they execute the tasks during usability test. This will generally exclude an interaction between the participant and the tester during the test that will affect the quantitative performance data.

#### *D. Co-Discovery Learning*

During a co-discovery learning [Nielsen (1993)] session, two participants attempt to perform the tasks together while the tester observes their interaction.

During a usability test, two test users attempt to perform tasks organized while being observed. They are to aid each other in the same manner as they would if they were working together to accomplish a common goal using the product.

#### *E. Question Asking Protocol*

This method [Dumas and Redish (1993)] is an extension of the thinking-aloud protocol wherein testers prompt participants by asking direct questions about the interface. The goal of such questioning is to enable the tester to get an even better understanding of the participant's mental model of the system.

#### *F. Retrospective Testing*

This method [Nielsen (1993)] is a follow-up to any other videotaped testing session wherein the tester and participant review the videotape together. During this review, the tester asks the participant questions regarding her behavior during the test. The goal of this review is to collect additional information from the usability test.

#### *G. Teaching Method*

For this method [Vora and Helander (1995)], the participant interacts with the system first to develop expertise to subsequently teach a novice user about the system. The novice user serves as a student and does not enthusiastically participate in problem solving. The participant does the problem solving, describes to the novice user how the system works, and shows a set of predetermined tasks.

#### *H. Thinking Aloud Protocol*

The Thinking-aloud protocol [Nielsen (1993)] requires participants to articulate their thoughts, feelings, and opinions during a usability test. One goal of this approach is to empower the tester to get a better understanding of the participant's mental model during interaction with the interface.

It encourages users to express out loud what they are looking at, thinking, doing, and feeling, as they perform tasks

#### *I. Shadowing Method*

Shadowing is an alternative to the thinking-aloud protocol wherein an expert user sits next to the tester and explains the participant's behavior during the testing session. Evaluators use this method in situations where it is inappropriate for participants to think aloud or talk to the tester.

During a usability test, the tester has an expert user (in the task domain) sit next to him/her and explain the test user's behavior to the tester. This technique is used when it's not appropriate for the test user to think aloud or talk to the tester while working on the tasks

#### IV. INQUIRY METHODS

Inquiry methods entail feedback from users and are often employed during usability testing. However, the focus is not on studying specific tasks or measuring performance. Rather the goal of these methods is to collect subjective impressions about innumerable aspects of a UI.

##### A. Field Observation

Field observation (Hom, 2003) is a field research method that in which product develop team member visits the user at the user's work place, observe the user's work activities; accumulate artifacts or gather data about the physical traits that marks the work place by photographing, note taking, or sketches; and interview the user about their work.

Human factors engineers go to representative users's workplace and observe them work, to understand how the users are using the system to accomplish their tasks and what kind of mental model the users have about the system. This method can be used in the test and deployment stages of the development of the product

##### B. Focus Groups

Focus group (Rosenbaum et al., 2002) originated as a market research method is a form of data gathering. A focus group is a meeting of about six to nine users wherein users discuss issues relating to the system. It is normally used during the product conceptualization phase in the product cycle where potential users are asked for their opinion on a potential product. The evaluator plays the role of the moderator and accumulates the desirable information from the discussion.

Its main disadvantage is that it does not test the actual user interaction with the system.

##### C. Interviews

An interview [Nielsen (1993)] is essentially a discussion session between a single user and an interviewer. During an interview, an evaluator enquires a user a sequence of questions about system issues to guide the discussion.

In an evaluation interview, an interviewer reads the questions to the user, the user replies verbally, and the interviewer records those responses. The methods of interviewing include unstructured interviewing and structured interviewing.

Unstructured interviewing methods are used during the earlier stages of usability evaluation. The objective of the investigator at this stage is to gather as much information as possible concerning the user's experience. The interviewer does not have a well-defined agenda and is not concerned with any specific aspects of the system. The primary objective is to

obtain information on procedures adopted by users and on their expectations of the system.

Structured interviewing has a specific, predetermined agenda with specific questions to guide and direct the interview. Structured interviewing is more of an interrogation than unstructured interviewing, which is closer to a conversation.

##### D. Logging Actual Use

It comprises automatic collection of statistics by the computer about the detailed use of the system. Typically an interface log contains statistics about the frequency with which the user has used each feature and frequency of various events e.g. error messages, undo, redo, etc. [Nielsen (1993)].

##### E. Questionnaires

A questionnaire [Soken (1993)] is a measurement tool designed to assess a user's subjective contentment with an interface. It is a list of questions that are circulated to users for responses. Responses on a questionnaire are usually quantitative.

Logging involves having the computer automatically collect statistics about the detailed use of the system. It is useful because it shows how users perform their actual work and because it is easy to automatically collect data from a large number of users working under different circumstances

##### F. Surveys

Survey is a widespread method to send out inquires and collect data from a large population in a short period of time. During a survey, an evaluator asks a user pre-determined questions and records responses. They could be done over the telephone, in person, over the mail or email.

#### V. CONCLUSIONS

Though categories of UEMs are becoming somewhat well-defined in the HCI discipline, methodologies for evaluating and comparing UEM effectiveness are not yet well-established. We believe it is possible to develop stable and consistent criteria for UEM effectiveness.

It was our objective in this paper to aid in lighten the problems of variation, incompleteness, and inconsistency in UEM evaluate and comparison studies. We urge cautious consideration of comparison criteria, both by researchers who perform UEM evaluation and comparison studies and by practitioners who use those studies to understand the relative merits of particular UEMs.

#### REFERENCES

- [1] Seffah, A., Donyaee, M., kline, R. B., & Padda, H. K. (2006). Usability measurement and metrics: A consolidated model. *Software Quality Journal*, 14, 159-178
- [2] Virzi RA, Sorce JF, Herbert LB. A comparison of three usability evaluation methods: heuristic, think-aloud, and performance testing. *Proceeding of the Human Factors and Ergonomics Society 37th Annual Meeting*. 1993:309-313.
- [3] J. Nielsen "Usability Engineering" p.198, Academic Press, 1993.

- [4] N. Soken, B. Reinhart, P. Vora, & S. Metz "Methods for Evaluating Usability" (Section 5B), Honeywell, Dec. 1993.
- [5] P. Vora & M. Helander "A Teaching method as an alternative to the concurrent think-aloud method for usability testing", in Y. Anzai, K. Ogawa & H. Mori "Symbiosis of Human and Artifact", pp.375-380, 1995.
- [6] "An empirical study of perspective-based usability inspection", by Zhijun Zhang, Victor Basili, and Ben Shneiderman. *Proceedings of the Human Factors and Ergonomics Society 42nd Annual Meeting*. pp. 1346-1350. Chicago, 1998.
- [7] Quesenbery, W. (2001). What does usability mean: Looking beyond "ease of use." *Proceedings of the 18th Annual Conference Society for Technical Communications*.
- [8] Abran, A., Khelifi, A., & Suryn, W. (2003). Usability meanings and interpretations in ISO standards. *Software Quality Journal*, 11, 325-338
- [9] Dix, Finley, J., Abowd, G. and Beale, R., *Human-Computer Interaction*, 2nd ed. Prentice-Hall, 1998.